

## Instruction

Pick up the bread near the cabinet

## Video



## VINE (ours)

### Concept Extraction

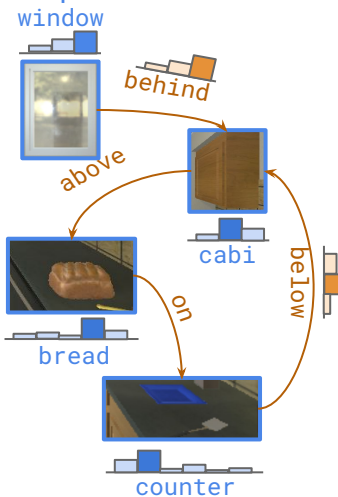


entities:  
[bread, cabinet]  
attributes:  
[yellow]  
relations: [below]

### Entity Identification



### Probabilistic Scene Graph



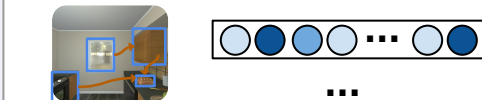
## Augmented Instruction

Pick up the bread near the cabinet. The bread is at (300, 350, 400, 450), the bottom right corner of the image, on the counter.

## Augmented Video



## Augmented Embedding



## MLLM with SG

**The bread is located on the right side of the image**, slightly towards the front, below the cabinet.

*Neurips 2025 Spotlight*



## VLA with SG

**Action chunk:**  
base\_vel=[+0.22,+0.00 -0.05]  
lift\_pos=0.11  
armSE3 = [0.61, -0.08, 0.92 |  
0, 1.57, 0]  
gripped\_command=CLOSE(0.4)

*CVPR 2026 submission*